



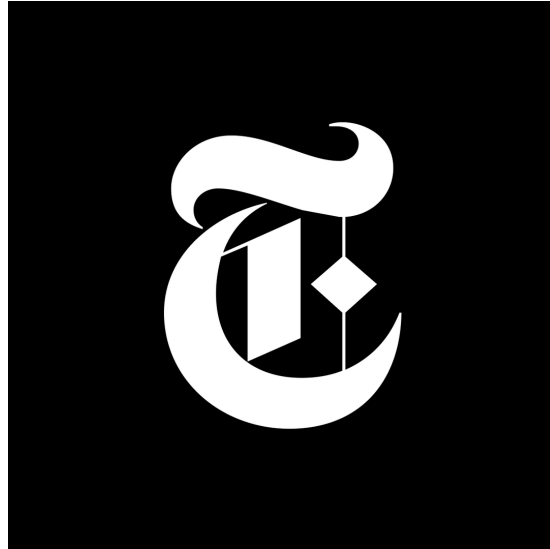
鲸析

Project Proposal

2022

[ZG104.GITHUB.IO](https://zg104.github.io)

DATA SOURCE



COVID-19 Google Public Data

The New York Times is releasing a series of data files with cumulative counts of coronavirus cases in the United States, at the state and county level, over time.

They are compiling this time series data from state and local governments and health departments in an attempt to provide a complete record of the ongoing outbreak.

Licence: <https://github.com/nytimes/covid-19-data/blob/master/LICENSE>

ABOUT DATA

INTRO

COVID-19 has been rampant for 2 years and is still mutating.

The data began with the first reported coronavirus case in Washington State on Jan. 21, 2020.

DATA

Historical State-Level Data

```
date,state,fips,cases,deaths
2020-01-21,Washington,53,1,0
...
```

The New York Times is releasing a series of data files with cumulative counts of coronavirus cases in the United States, at the state and county level, over time.

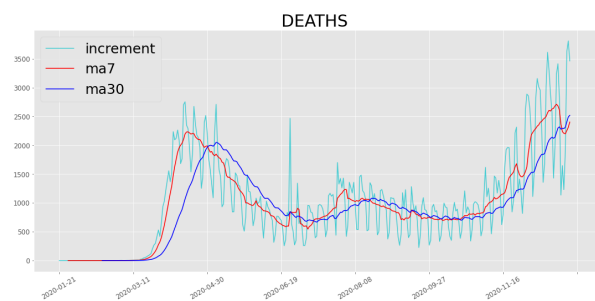
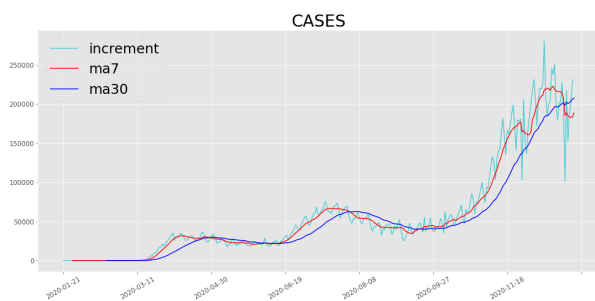
METADATA

	explanation	variable	dtype	keep	unique	missing_values
columns						
date	date(format:YYYY-MM-DD)	object	object	True	795	0(0.0%)
state	state	object	object	True	56	0(0.0%)
fips	fips	ordinal	int64	True	56	0(0.0%)
cases	cumulative cases	ordinal	int64	True	ignore	0(0.0%)
deaths	cumulative deaths	ordinal	int64	True	ignore	0(0.0%)

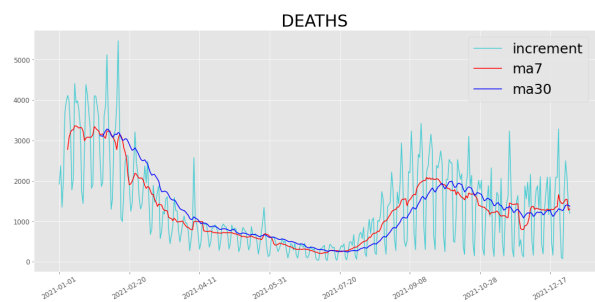
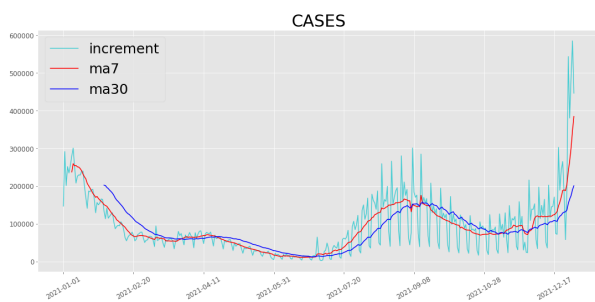
OVERVIEW

The spread and outbreak of Delta virus has once again brought the United States into a deadlock. Through the epidemic data of various counties and states from the beginning of 2021 to the present, I would like to do some descriptive analysis on it, so as to analyze the trend of spread through these time series data. Predicting the future trend of the epidemic using ARIMA method

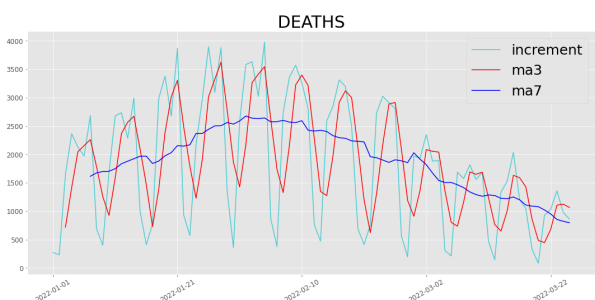
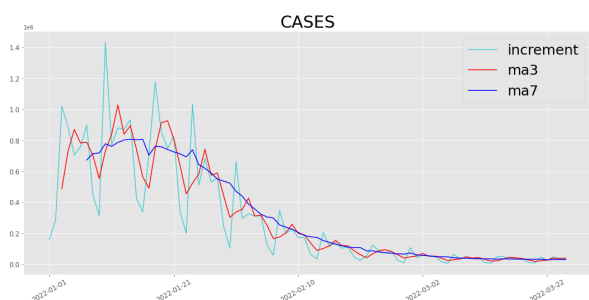
TIME RANGE: 2020-2021



TIME RANGE: 2021-2022

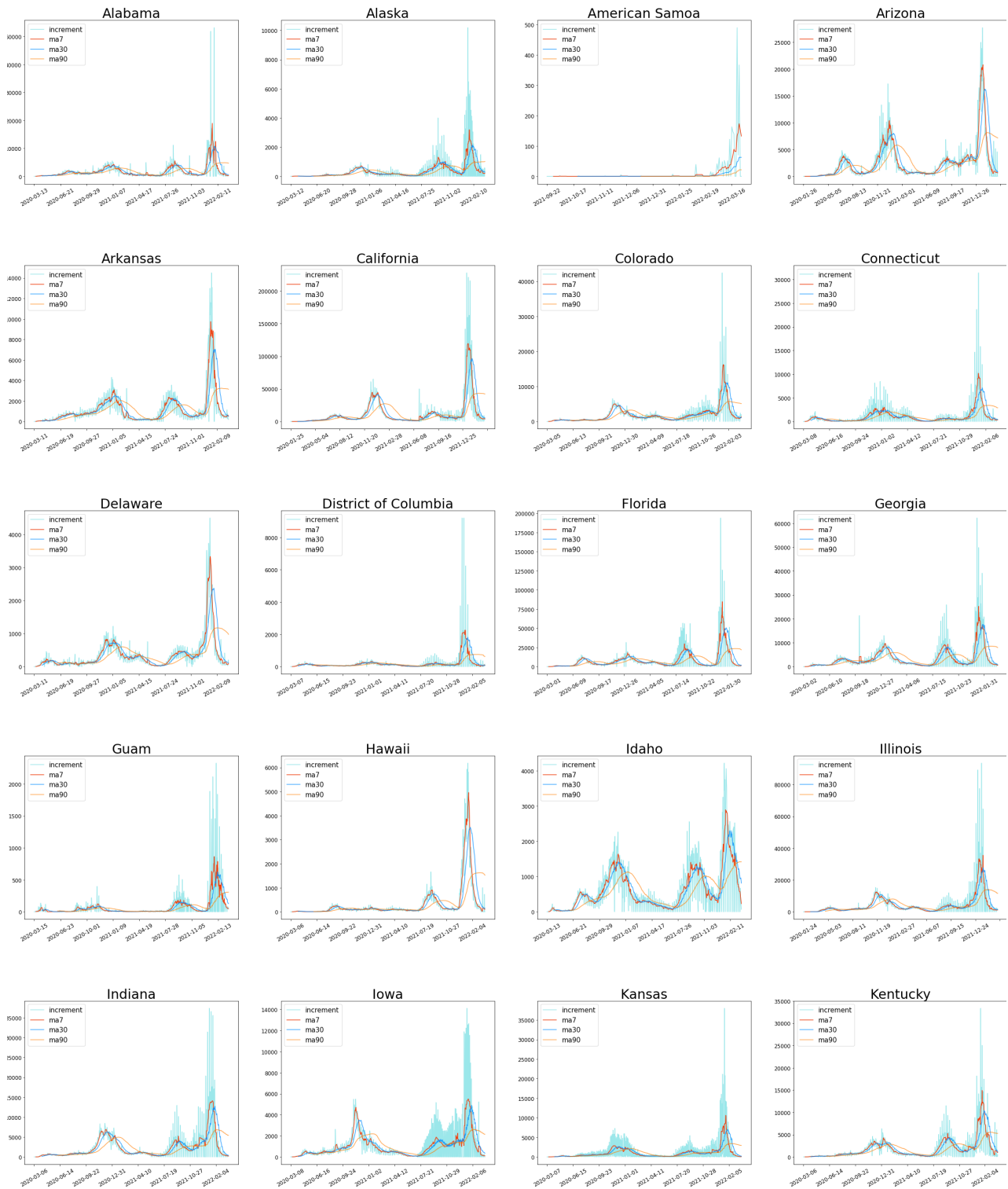


TIME RANGE: 2022-PRESENT



OVERVIEW

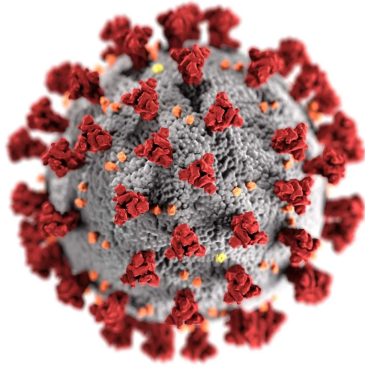
ACROSS THE STATES



FULL IMAGE LINK

<https://gitee.com/gzjzg/WhaleDataScienceProject/raw/master/pic/COVID19AcrossState.png>

EXPECTATION



Begin by elaborating in more detail, I will take these necessary steps to dive in:

- Create a map using FIPS to present the severity of all counties or states in the USA
- Calculate common statistics in analysing the covid spread situation in each states
- Perform data prediction using ARIMA model by calculating the 7 day moving average

Since the upward trend of spreading is not going to diminish, I will update the data once a week. And I will keep the data for now as a training set so as to train the model, and test on the data updated in the future. Hope I could predict the trend pattern in order to help research the spread speed.